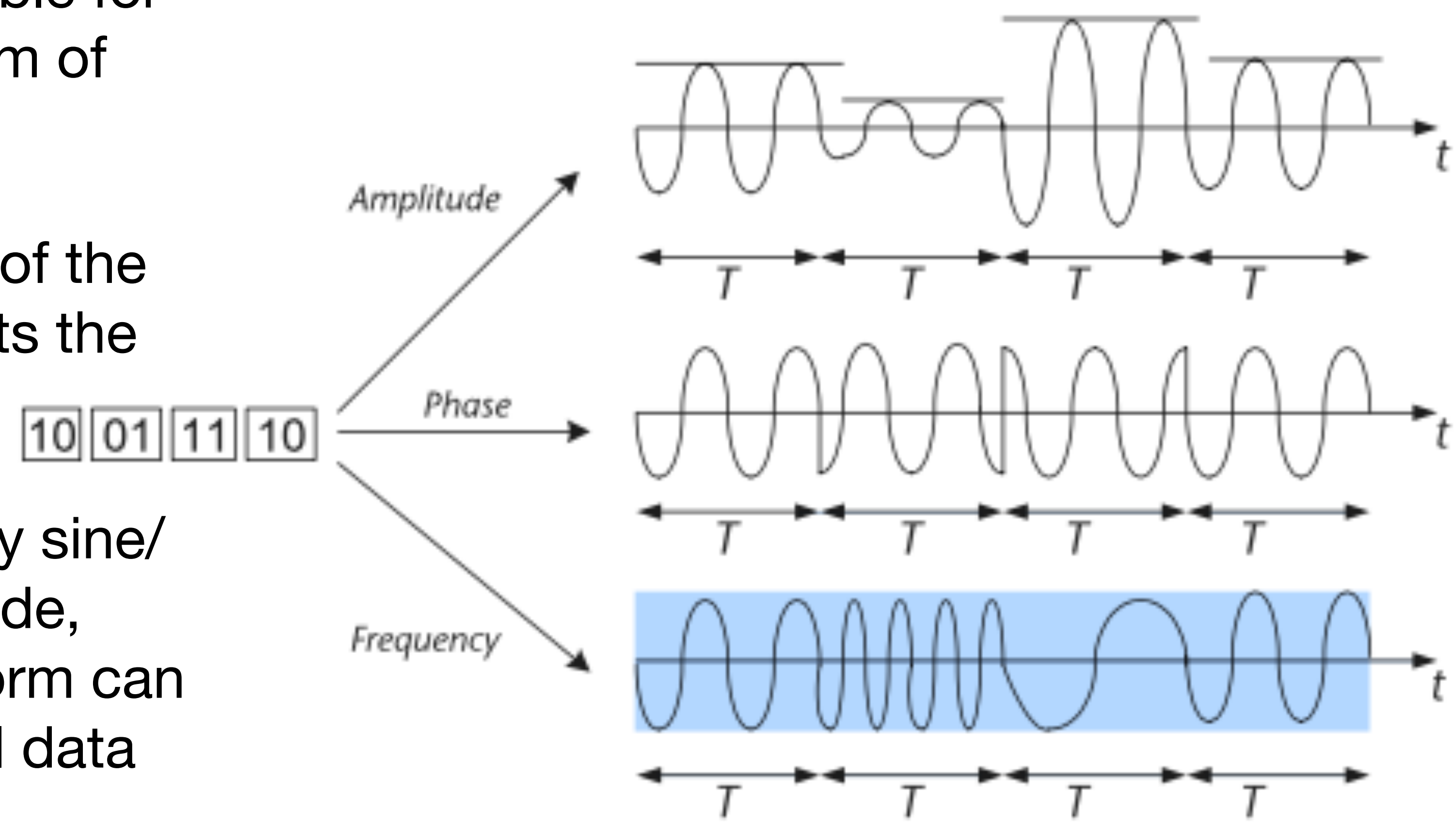
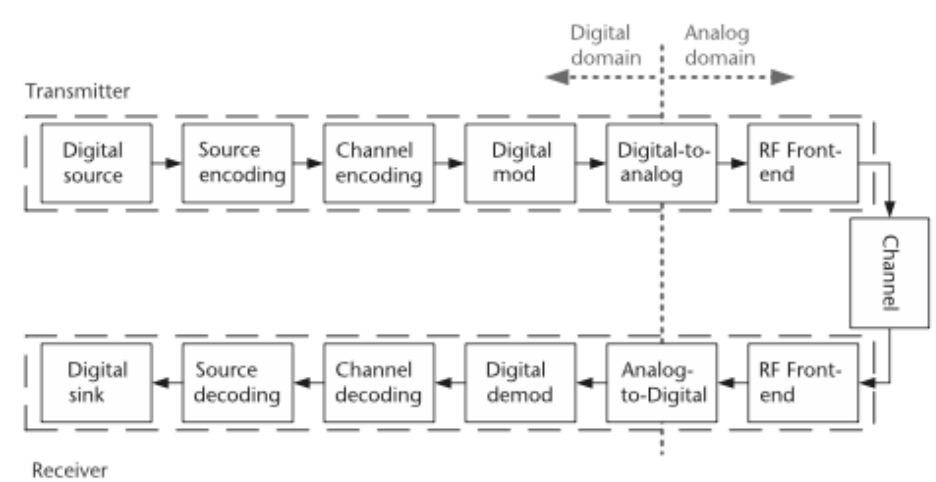
EEU4C21/CSP55031/EEP55C26:
Open Reconfigurable Networks

Modulation, Encoding and Detection

Digital transmission

- A digital transceiver is responsible for the translation between a stream of digital data and EM waveforms
    - The physical characteristics of the waveform uniquely represents the digital information (in bits)
- Since EM waveforms are mostly sine/cosine, parameters like amplitude, phase, frequency of the waveform can be used to uniquely map digital data per time interval T
- There is more to a transceiver than this mapping - the blocks within the system aid in reliable communication across the non-ideal channel





Source encoding

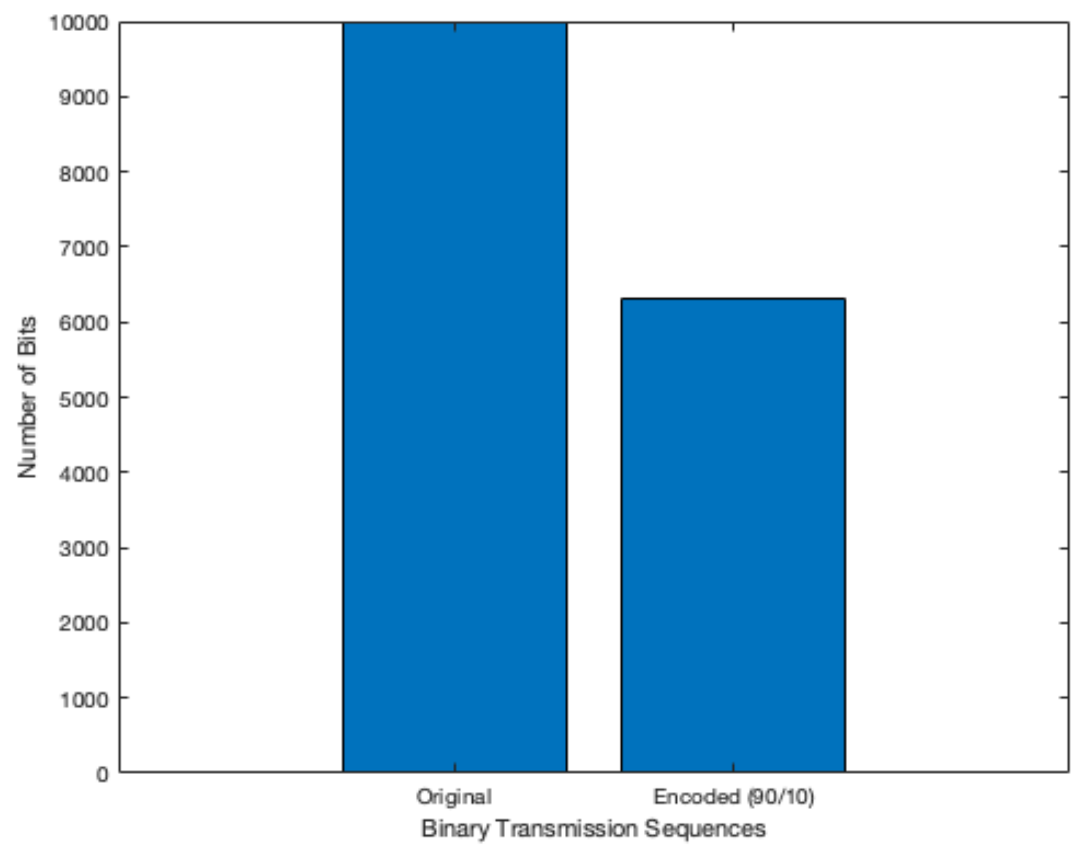
Thee shalt not wasteth bandwidth!

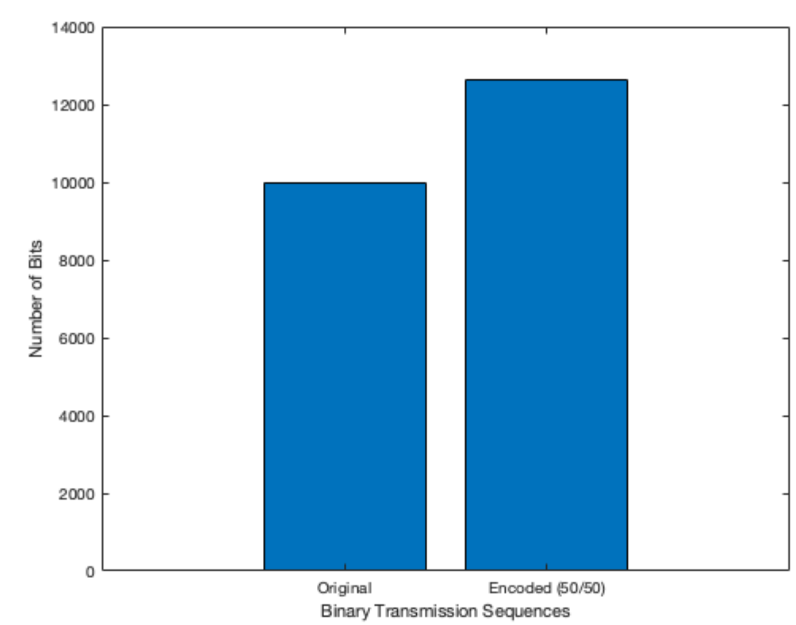
- Most source data has some level of redundancy
- Removing redundant information reduces the amount of information to be transmitted
    - Reduces time, computations, power…
- Source encoding takes source symbols $\underline{u}$ and maps them to a encoded set of symbols $\underline{v}$
    - such that a sample $v_i \epsilon \underline{v}$ is as close to random as possible and elements within $\underline{v}$ are unrelated (or uncorrelated)

Source encoding

Consider the following case:

- Source 1 is a binary vector with 90% of values at level ‘1’ and 10% at level ‘0’
    - Use a mapping scheme that takes all continuous strings of ‘1’ and replace it with a binary equivalent of the length of the strings
        - i.e., something like 1111_1111_1111_0 -> 1100_0
- What happens if source 2 has equal number of 0’s and 1's?





Channel encoding

- Transmission channel is non-ideal and introduces errors
- Channel encoding is design to correct for channel transmission errors by introducing controlled redundancy into the data transmission
    - Recall that source encoding removes redundancy in the source data - typically, random in nature
    - Controlled redundancy integrates a structured redundancy that is known at both txr and rxr, and is designed to combat effects of bit errors in transmission
- Each source encoded vector of length K i.e., $\forall\, v_l \,\epsilon\, v,$ where $l = 1,2,...2^K$, is assigned a unique codeword $c_l \,\epsilon\, \mathbb{C}$ of length N
    - $N - K = r$ is the number of controlled bits of redundancy added
- $\mathbb{C}$ us called codebook, and code rate is ratio of number of information bits $\{k\}$ to the size of the codeword $\{N\}$

How good is your coding scheme?

- To determine the effectiveness of a set of codewords within a codebook, we can use <u>Hamming distance</u>
    - i.e., relative distance between two codewords c_i and c_j given by $d_H(c_i, c_j)$ = number of components (bits) in which c_i and c_j are different
    - minimum Hamming distance within a set of codewords determines its effectiveness
        - e.g., in a codeword {101, 111}, $d_{H,min} = 1$ whereas in {101, 010}, $d_{H,min} = 3$
- In the event of corruption during transmission, decoding spheres can aid in reception - map the received symbol to the nearest value in the codeword
    - see example on page 124 for a MATLAB example using 1/3 repetition code

Channel capacity

- Claude Shannon's theorem gives an upper limit on the data rate for a specific transceiver and channel combination for error-free transmission
  - It states that there exists a code rate $R_c = k/N < C,$ such that $N \to \infty \implies P_e \to 0$
  - Conversely, no R_c exists for $R_c \geq C$ i.e., C is the absolute capacity limit without causing errors
  - Or in terms of received SNR and transmission bandwidth, $C = B\log_2(1 + \text{SNR}) \qquad [b/s]$
  - i.e., the maximum achievable data rate in relation to information capacity
    - Other uses include trade-off analysis between B and SNR, compare noise performance of modulation schemes

Digital Modulation

- Digital message signal modulates a <u>continuous waveform</u>
  - Modifying the amplitude, phase or frequency of the signal during each <u>symbol</u> period, T
  - Symbol is typically a collection of b bits forming a binary message m_b, which is then mapped to a symbol
    - i.e., each possible 2^b value of m_b, results in a unique signal $s_i(t), 1 \leq i \leq 2^b$ that can be used to modulate the continuous waveform

An aside: Power efficiency

Power efficiency is used to compare efficiency of modulation schemes and symbol mapping (i.e., bit to symbol in terms of transmit power per symbol)

We know that energy of a symbol is $E_s = \int_0^T s^2(t)dt$

For a given set of M symbols, each with probability of occurrence given by $P(\cdot)$, then for the modulation scheme,

$$\bar{E}_s = P(s_1(t)) \cdot \int_0^T s_1^2(t) + \ldots + P(s_M(t)) \cdot \int_0^T s_M^2(td)dt$$

To normalise this across different schemes, we divide by number of bits per symbol $b = log_2(M)$, yielding average energy per bit

$$\bar{E}_b = \frac{\bar{E}_s}{b} = \frac{\bar{E}_s}{log_2(M)}$$

An aside: Power efficiency

Also, similarity between two symbols can be measured by the *Euclidean distance*

$$d_{ij}^2 = \int_0^T (s_i(t) - s_j(t))^2 dt = E_{\Delta s_{ij}}, \text{ where } \Delta s_{ij} = s_i(t) - s_j(t)$$

For comparison, we consider the worst case scenario, where the distance is minimum

$$d_{min}^2 = \min_{s(i)(t), s_j(t), i \neq j} \int_0^T (s_i(t) - s_j(t))^2 dt$$

Hence, power efficiency of the set of signals used for a modulation scheme is given by

$$\epsilon_p = \frac{d_{min}^2}{\bar{E}_b}$$

Modulation schemes

Pulse Amplitude Modulation

- Message information is encoded in the amplitude of a series of signal pulses

- Simplest form is binary PAM (B-PAM) - maps bits to a waveform $s(t)$ with two amplitude levels; for e.g.,

$$\text{"1"} \rightarrow s_1(t) \;\&\; \text{"0"} \rightarrow s_2(t)$$

- $s(t)$ is defined across the time period T and zero otherwise; however, since for B-PAM symbol time = duration of bits, the rate of transmission $R_b = 1/T$ bps

- In the simplest case, s(t) can be a rectangular pulse where $s_1(t) = s(t) = A \cdot [u(t) - u(t-T)]$, and $s_2(t) = -s(t)$
  - i.e., transmit +A if the digital signal is “1” for a time window T, transmit -A if digital signal is “0”.

We can derive $\bar{E}_b = A^2T$, $d_{min}^2 = 4A^2T$ and $\epsilon_p = 4$ for B-PAM

Modulation schemes

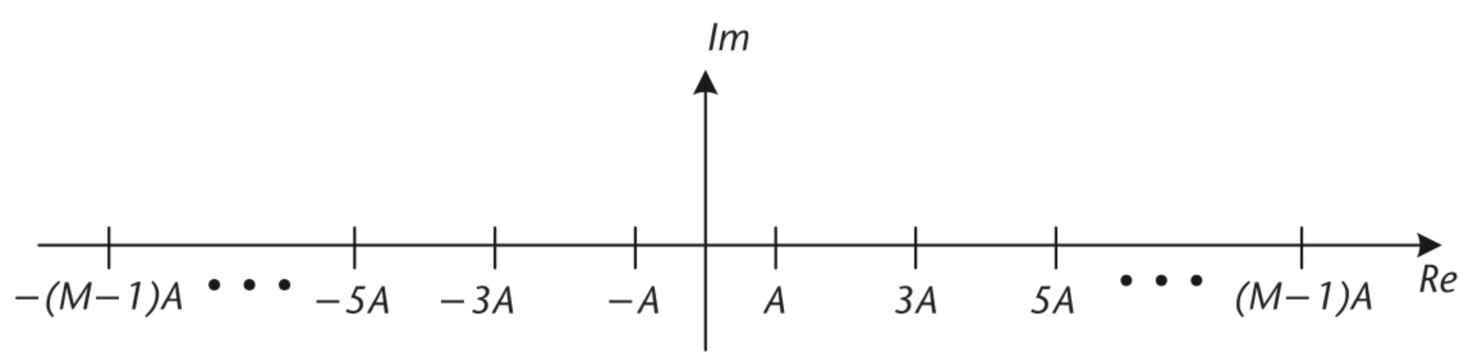
***M*-ary Pulse Amplitude Modulation**

- A generalised form of B-PAM, where we map *M* binary sequences to *M* possible unique amplitude levels

- Using the rectangular pulse as the pulse shape, we can express M-PAM as

$$s_i(t) = A_i \cdot p(t), \text{ for } i = 1,2,\ldots,M/2$$
$$\text{where, } A_i = A(2i-1) \text{ and } p(t) = u(t) - u(t-T)$$

- i.e., depending on the message m_i the transmitted pulse with have an odd multiple of $\pm A$



We can derive $\bar{E}_b = \dfrac{A^2T(2^{2b}-1)}{3b}$, $d^2_{min} = 4A^2T$ and $\epsilon_{p,M-PAM} = \dfrac{12b}{2^{2b}-1}$ for M-PAM

Modulation schemes

Quadrature Amplitude Modulation

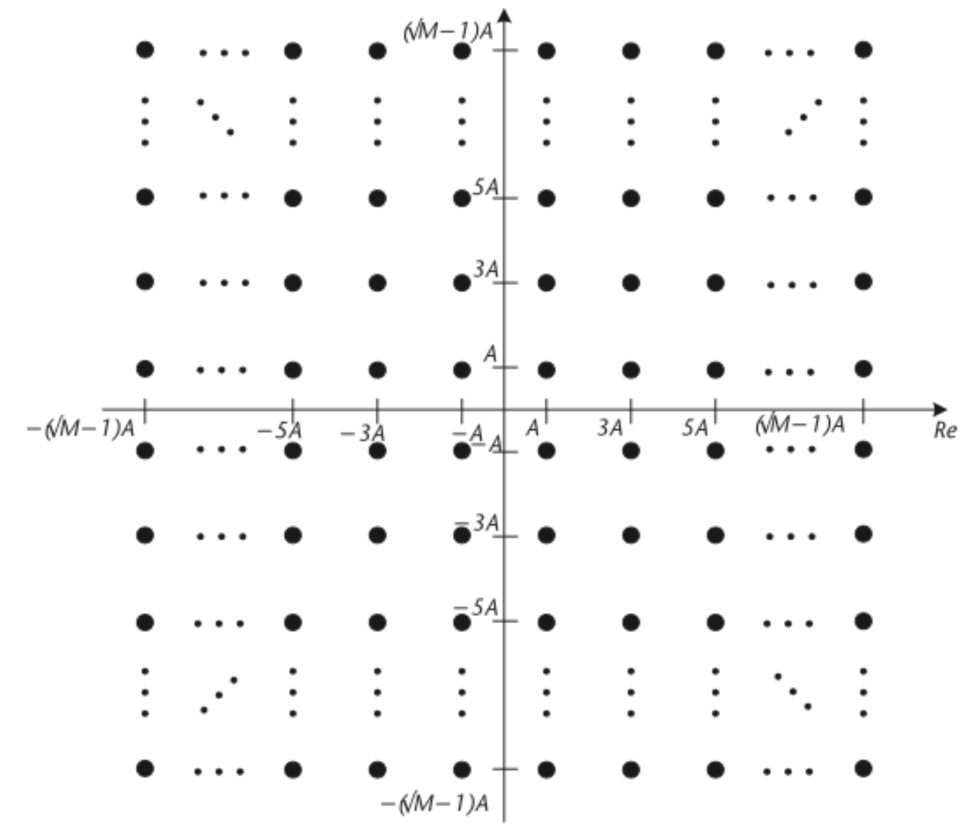
- What if we can use the imaginary axis to also represent information?
  - Two-dimensional scheme using in-phase and quadrature components (i.e., sine/cosine) as modulating waveforms → double the transmission rate; pack 2-bits per symbol

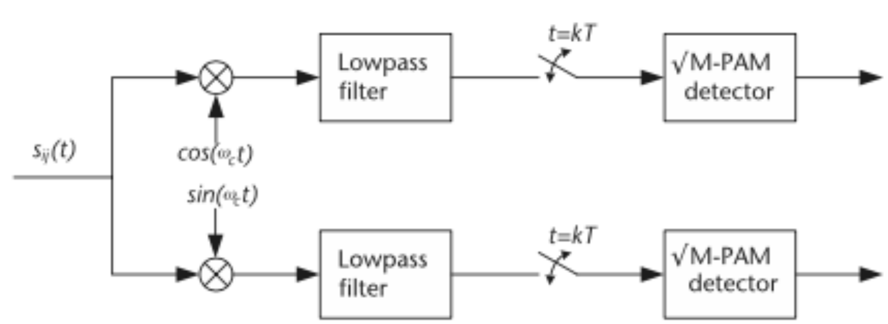
"00" → $s_{00}(t)$, "01" → $s_{01}(t)$, "10" → $s_{10}(t)$, "11" → $s_{11}(t)$,

- Think of rectangular QAM as two orthogonal PAM
  - constellation consisting of M unique waveforms $\Longrightarrow$ two $\sqrt{M}$-PAM transmissions in orthogonal dimensions

i.e.,$s_{ij}(t) = A_i \cdot \cos(\omega_c t) + B_j \cdot \sin(\omega_c t)$

  - Hence, this results in a 2-D constellation with a straightforward receiver architecture





We can derive $\bar{E}_b = A^2 T \dfrac{2^b - 1}{3b}$, $d^2_{min} = 2A^2T$ and $\epsilon_{p,M-QAM} = \dfrac{3! \times b}{2^b - 1}$

Modulation schemes

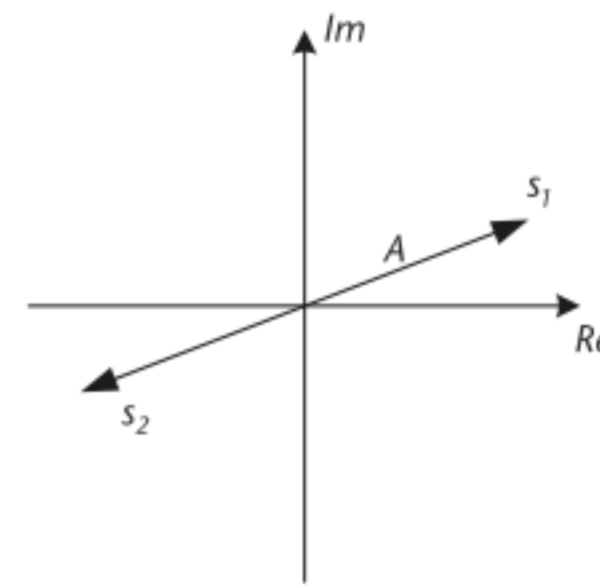
Phase Shift Keying

- Message information is encoded by modulating the phase of a reference signal
- Finite number of phases of a signal ascend to unique patterns of binary digits - each phase encodes an equal number of bits (i.e., a symbol)
- Demodulator determines the phase of the received signal and maps it back to the symbol it represents - i.e., receiver must know the phase of the reference signal to compare — or **coherent detection**
- Mathematically, we can define PSK as

$$s_i(t) = A\ \cos(2\pi f_c t + (2i-1)\frac{\pi}{m}),\ \text{for } i = 1,\ldots,\log_2 m$$

A is constant amplitude , $f_c \rightarrow$ carrier frequency

$(2i-1)\dfrac{\pi}{m}$ is phase offset for each symbol



- As with B-PAM, simplest case is when bits are directly encoded — i.e., binary PSK (B-PSK)

$$\text{"1"} \rightarrow s_1(t) = A \cdot \cos(\omega_c t + \theta)$$
$$\text{"0"} \rightarrow s_2(t) = A \cdot \cos(\omega_c t + theta + \pi) = -\, s_1(t)$$

  - i.e., transmit $A\cos\left(\omega_c t + \theta\right)$ if the digital signal is “1” for a time window T, transmit $-A\cos(\omega_c t + \theta)$ if digital signal is “0”

We can derive $\bar{E}_b = \dfrac{A^2 T}{2}$, $d^2_{min} = 2A^2 T$ and $\epsilon_{p,BPSK} = 4$

Modulation schemes

Quadrature Phase Shift Keying

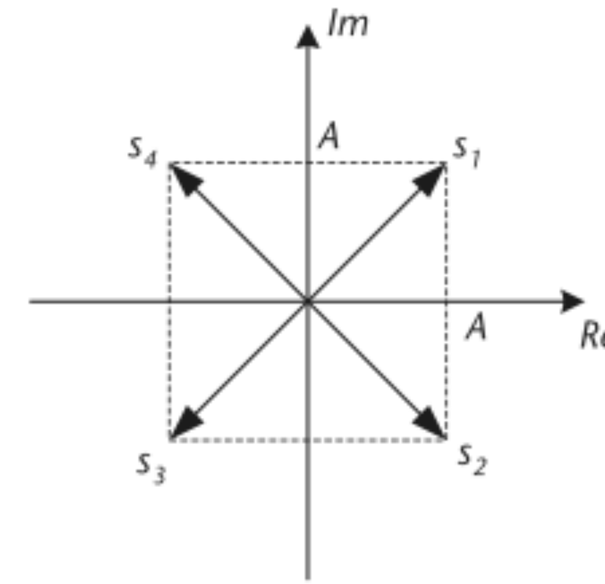
- Like with PAM, we can use four distinct waveforms per modulation scheme to arrive at QPSK

- Mathematically, we can define QPSK signal waveform as

$$s_i(t) = \pm A \cdot cos(\omega_c t + \theta) \pm A \cdot \sin(\omega_c t + \theta)$$

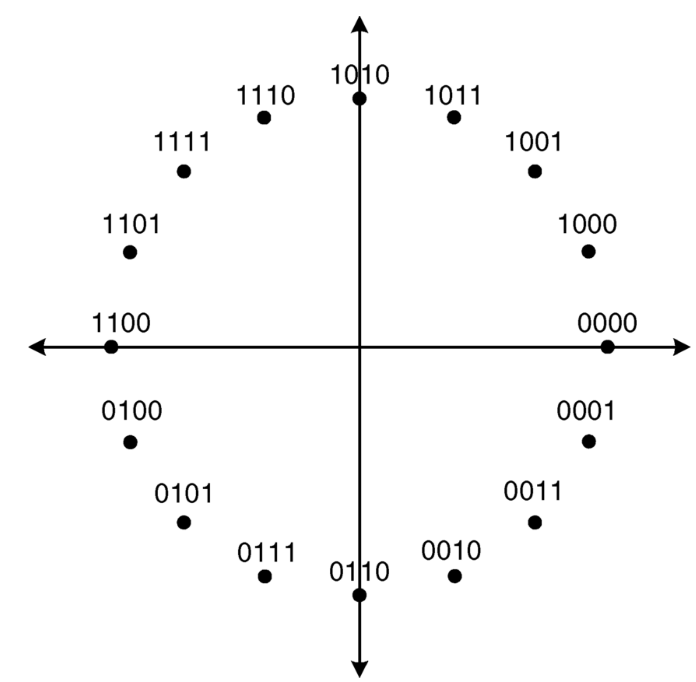
  - i.e., each signal has same amplitude (A) but one of the 4 possible phase values (i.e., 2 bits per symbol)

  - NOTE that QPSK has the same power efficiency as BPSK, but with 2-bits per symbol, making it more efficient



Finally, we have the general case of M possible phase values creating a signal constellation of M equally spaced points on a circle: 16-PSK in the figure

$$s_i(t) = A \cdot \cos\left(\omega_c t + \frac{2\pi i}{M}\right), \text{ for } i = 0,1,\ldots,M-1$$



We can derive $\bar{E}_b = \dfrac{A^2T}{2}$, $d^2_{min} = 2A^2T$ and $\epsilon_{p,QPSK} = 4$

Modulation schemes

Summary of modulation schemes

- To determine the trade-offs between different schemes, we compare it relative to QPSK as $\delta\text{SNR} = 10 \cdot \log_{10}\left(\frac{\epsilon_{P,QPSK}}{\epsilon_{P,\text{other}}}\right)$

Table 4.1 δSNR Values of Various Modulation Schemes

M	b	M-ASK	M-PSK	M-QAM
2	1	0	0	0
4	2	4	0	0
8	3	8.45	3.5	—
16	4	13.27	8.17	4.0
32	5	18.34	13.41	—
64	6	24.4	18.4	8.45

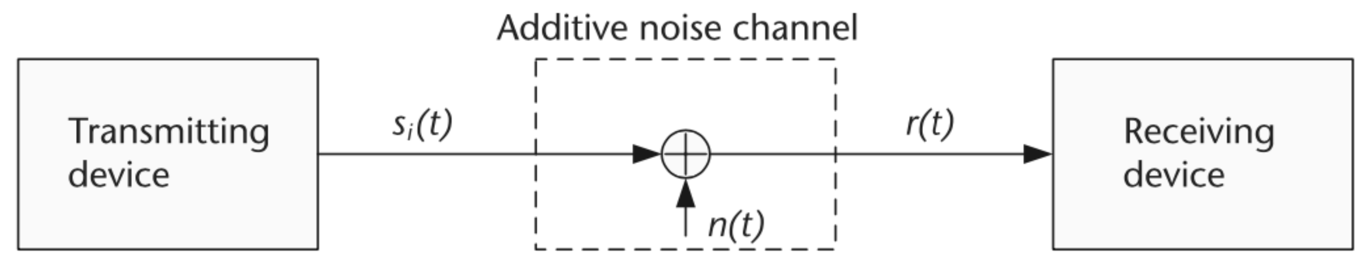
- In general, two-dimensional modulation (Quadrature variants) perform better than the one-dimensional schemes

- All schemes here are linear $\Longrightarrow$ similar levels or receiver complexity

- Another comparable factor is error performance (measured with BER) - here, for the same amount of noise, QPSK performs worser than 4-PAM and 4-QAM (see code 4.6 in textbook and 4.7 for waterfall curves using Monte Carlo simulations)

Optimal Detection

- Decision theory or signal detection theory is used to discern between signal & noise



- Receiver only observes $r(t)$, corrupted version of transmitted signal $s_i(t)$ by noise signal $n(t)$

- Detection problem is thus:

  Given $r(t), 0 \leq t \leq T$,
  determine which $s_i(t), i = 1,2,\ldots M$
  was transmitted

We know that:

$$E\{n_k\} = E\left\{ \int_0^T n(t)\phi_k(t)dt \right\} = 0$$

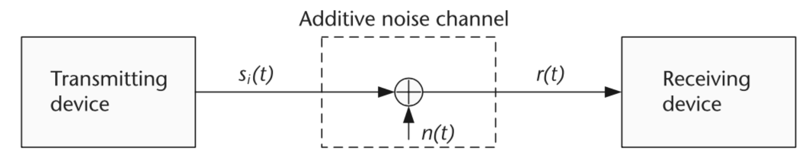
$$E\{n_k n_l\} = \frac{N_0}{2}\delta(k-l) = \frac{N_0}{2} = \sigma^2$$

$$p(n) = p(n_1, n_2, \ldots, n_N) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-||n||^2/2\sigma^2}$$

i.e., Gaussian noise conditions

Note: $\phi_k(t)$ is the orthonormal function corresponding to n_k, see Gram-Schmidt Orthogonalisation & signal-space representation

Optimal Detection: Aside



Decision rules

With the pdf of noise known, we can define probabilistic rules to estimate $\hat{m}_k$, given **r** was received with some noise

Minimise P(error) or P(e) $\rightarrow P(\hat{m}_i \neq m_i$

Maximise P(correct) or P(c) $\rightarrow P(\hat{m}_i = m_i)$

Assuming that r $= \rho$ was received, i.e.,$\rho = s_k +$ n $\rightarrow \hat{m} = m_k$,

Maximising P(c) becomes

Maximise $P(c \,|\, r = \rho) \implies P(s_k \,|\, \rho) \geq P(s_i \,|\, \rho) \forall i, i \neq k$

Through conditional probability & Bayes rule, this is simplified as

$$\max_{s_i} P(s_i \,|\, r = \rho) = \max_{s_i} \frac{p(\rho \,|\, s_i) P(s_i)}{p(\rho)} = \max_{s_i} p(\rho \,|\, s_i) P(s_i) \forall i$$

This expression can yield two detectors: MAP detector (in the current form) and Maximum likelihood detector, when s_i are equiprobable.

MAP - *maximum a posteriori* decision rule

Max Likelihood detector

Using the pdf of the noise signal and further simplification using logarithms, the decision rule in case of Max likelihood detector boils down to

$$\max_{s_i} \ln(p(\rho \mid s_i)) = \min_{s_i} ||\rho - s_i||$$

or $s_k = \arg\min_{s_i} ||\rho - s_i|| \rightarrow \hat{\text{m}} = \text{m}$

i.e., choose the symbol vector which has minimal distance from the received vector in the symbol-space

This could be implemented either using a matched filter or correlation detector

- Matched filter: filter coefficients match the energy of the received signal with different symbols and chooses the one with maximum similarity - requires knowledge about waveforms and noise characteristics

- Correlation detector: correlate received signal with different possible symbols, choose the one with maximal correlation (post normalisation) - only knowledge about waveforms need to be known [requires perfect synchronisation]

Matched Filter v/s Correlation detector